

Superiorization of EM Algorithm and Its Application in Single-Photon Emission Computed Tomography(SPECT)

Shousheng Luo

School of Mathematical Sciences, Peking University, Beijing 100871, China

Tie Zhou

School of Mathematical Sciences, Peking University, Beijing 100871, China

Abstract

In this paper, we presented an efficient algorithm to implement the regularization reconstruction of SPECT. Image reconstruction with priori assumptions is usually modeled as a constrained optimization problem. However, there is no efficient approaches to solve it due to the ill-posedness and large scale of the problem. In this paper, we use the superiorized approach of the expectation maximization (EM) iteration to implement the regularization reconstruction of SPECT. We first investigated the convergent conditions of the EM iteration in the presence of perturbations. Then we illustrated the superiorized EM algorithm based on the convergent conditions, and proposed a modified version of it. Furthermore, we gave two concrete methods to generate perturbations for two special objective functions. Numerical simulations for SPECT reconstruction were conducted to validate the performance of the proposed algorithms. The simulations show that the superiorized EM algorithms are more stable and robust for noise pollution and selection of initial point than the classic EM algorithm, and the reconstructed images by the proposed algorithms outperform the classic EM algorithm in terms mean square error and visual quality.

Keywords: EM algorithm, superiorization, SPECT.

1 Introduction

Single-photon emission computed tomography (SPECT), which can visualize the physiological information of various organs with the help of radiopharmaceuticals [1, 2, 3]. The gamma-rays emitted by the injected radioactive material are recorded by a gamma camera rotated around the patient. The goal of SPECT is to reconstruct the radionuclide distribution from the measurements numerically. The variety of existing SPECT algorithms can be split into a family of analytical methods [4, 5, 6, 7, 8] and a wide class of iterative techniques [9, 10, 11].

From the analytic point of view, the SPECT reconstruction problem [3, 4, 7] is to inverse the attenuated Radon transform(aRt) of f (distribution of radiopharmaceutical)

$$R_a f(s, \varphi) = \int_R f(s\theta + t\theta^\perp) e^{-\int_t^\infty \mu(s\theta + \tau\theta^\perp) d\tau} dt, \quad (1)$$

where μ is a known function, referred to as the attenuation map of gamma-rays, $\theta = (\cos \varphi, \sin \varphi)$ and $\theta^\perp = (-\sin \varphi, \cos \varphi)$. In practice, f is a function within compact support Ω . Therefore, the integrand in (1) is zero outside a bounded interval; this integral is written over $(-\infty, +\infty)$ for convenience.

For the iterative method, the SPECT reconstruction problem is modeled as the inversion of the following linear system [9],

$$Ax = b, \quad (2)$$

where the elements of the observed data $b = (b_1, b_2, \dots, b_M) \in \mathbb{R}^M$, the unknown image $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ and the system matrix $A = (a_{ij}) \in \mathbb{R}^{M \times N}$ are all nonnegative. The aim is to reconstruct the unknown x as an image from the projection data b via stable algorithms. A solution is not feasible with conventional methods directly because of the noisy projection data b , ill-posedness and large scale of the problem.

For iterative methods, the system matrix A not only can model the attenuation of photon, but also can fuse some realistic factors, such as photon scattering and camera blurring. Therefore, we focus on the iterative methods in this paper. The expectation maximization (EM) algorithm [9, 12] and the algebraic reconstruction technique (ART)[13, 14] are two widely used technologies in imaging science, due to their simplicity, efficiency

and performance.

In practice, the EM algorithm is more appropriate for emission tomography including SPECT. Firstly, the EM algorithm maintains the nonnegative constraint in the iteration procedure. Secondly, the EM algorithm is relatively robust against data inconsistencies introduced by Poisson noise, because it seeks to minimize the Kullback-Liebler(K-L) distance between the measured data b and the projection of the estimated image Ax , which is equivalent to maximize the likelihood of Poisson distribution.

For SPECT reconstruction, it is one of the main issues to estimate the radionuclide distribution from low-counts projection data. This issue occurs quite frequently because of practical constraints, such as imaging hardware and scanning geometry. Furthermore, the decrease of the counts can reduce acquisition time, radiation dose and imaging cost efficiently. However, this would result in the strong deterioration of the observed data and the under-determinacy ($m \ll n$) of the linear system (2). In these situations, the reconstructed images by the EM algorithm are usually dominated by various distortions [1, 15, 16, 17], because the EM algorithm accepts any solution which minimizes the K-L distance.

The qualities of the reconstructions can be improved by regularization reconstruction methods. The regularization reconstruction for SPECT is usually modeled as a large scale constrained optimization problem. However, to our knowledge, there is no optimal algorithm to solve this problem efficiently. In this paper, we resort to an emerging approach called superiorization [18] to implement the regularization reconstruction of SPECT.

The superiorization of iterative methods, which was first proposed by the authors of [16], is a relaxation technology for constrained optimal problem. The superiorized algorithm lies between the feasibility-seeking algorithms, which seek a feasible point in the constrained set, and the optimal algorithms, which seek the maximum(minimum) point of objective function in the constrained set. The aim of superiorization algorithm is to look for a superior point of the objective function instead of the optimal point or just a feasible point in the constrained set. The basic idea of superiorization is to do the feasibility-seeking iteration method with perturbations from the objective function.

The superiorization of ART-like algorithms has been studied and applied to the total variation regularization reconstruction of computed tomography(CT) [15]-[20]. The authors of [15, 16] first investigated the convergence of the two variants of ART under summable perturbations for consistent case. For inconsistent case, the authors of [20] proved the convergence of symmetric version of ART in the presence of summable perturbations of the iterates. The superiorization of the EM algorithm was firstly proposed in [21], and applied to bioluminescence tomography. However, to our knowledge, it is still an open problem about the convergence of the superiorized EM algorithm.

In this paper, we will first discuss the convergence of the EM algorithm in the presence of perturbation in Section 2. A so-called bounded perturbation resilient (BPR) property of ART is vital in proof of convergence for the perturbed version of ART. However, we cannot prove the BPR property of the EM iteration so far, because of the nonlinearity of the EM operator. Therefore, we investigate the convergence of the perturbed EM algorithm under the following assumptions. Firstly, the perturbations should maintain the positivity of iterations. Secondly, the perturbations should go to zero with the increase of iterates. Lastly, the perturbed EM iteration should maintain the gradual decrease of the K-L distance between the observed data and the projection of estimation by each iteration.

Based on the convergent conditions, we present the superiorized EM algorithm and its modified version in Section 3. Furthermore, practicable techniques are given to produce ideal perturbations for two special objective functions, TV [22] and l^1 -norm [23], widely used in imaging science. While the proposed algorithms are applicable to diverse inverse problems, in this paper we restrict ourselves to demonstrate its usefulness to the SPECT reconstruction problem. Numerical results for SPECT reconstruction are given in Section 4. And as expectation, the superiorization algorithms output superior image comparing with the classic EM algorithm in terms of MSE and visual quality. Some conclusions and discussions are given in Section 5.

2 Perturbations Resilience of EM Iteration

For the sake of reference, we first introduce some notations and assumptions. In this paper, an image x is described as a vector of length N with individual elements x_j , $j = 1, 2, \dots, N$. When it is necessary to refer to pixels in the context of a 2D image we use the double subscript form $x_{p,q}$, where

$$j = (q - 1)W + p, p = 1, 2, \dots, H, q = 1, 2, \dots, W, \quad (3)$$

and integers W and H are, respectively, the width and height of the 2D image array, which has a total number of pixels $N = W \times H$. Denote by \mathbb{R}_+^N the region $\{x \geq 0 | x \in \mathbb{R}^N\}$. For the system matrix, we assume $H_j = \sum_i a_{ij}$,

and denote by $d_i(x) = \sum_j a_{ij}x_j$ the i th component of Ax for convenience. Furthermore, we introduce

$$g_{ij}(x) = \frac{a_{ij}b_i}{d_i(x)}, \quad (4)$$

$$f_j(x) = \frac{1}{H_j} \sum_i \frac{a_{ij}b_i}{d_i(x)} = \sum_i g_{ij}(x). \quad (5)$$

Let P denote the EM operator, then the EM iteration $x^{k+1} = P(x^k)$ for the problem (2) is defined as

$$x_j^{k+1} = P_j(x^k) = \frac{x_j^k}{H_j} \cdot \sum_i \frac{b_i a_{ij}}{\sum_t a_{it} x_t^k} = x_j^k \cdot f_j(x^k), \quad (6)$$

with an initial point $x^0 > 0$. Similarly, the perturbed version of the EM iteration is defined as

$$x_j^{k+1} = P_j(x^k + \beta_k v^k) = \frac{y_j^k}{H_j} \cdot \sum_i \frac{b_i a_{ij}}{\sum_t a_{it} y_t^k} = y_j^k \cdot f_j(y^k), \quad (7)$$

where $y^k = x^k + \beta_k v^k$. Here, the vector v^k and number β_k represent the perturbed direction and length for the k th iteration, respectively. Hereafter, we call the EM iteration without perturbation as the classic EM iteration.

It has been established that the sequence $\{x^n\}$ generated by the classic EM iteration converges to the minimizer of the K-L distance $I_A^b(x)$ between b and Ax on \mathbb{R}_+^N , where $I_A^b(x)$ is defined as

$$I_A^b(x) = I(b, Ax) = \sum_{i=1}^M b_i \log \frac{b_i}{d_i(x)} - \sum_{i=1}^M (b_i - d_i(x)), \quad (8)$$

where $I(\cdot, \cdot)$ denote the K-L distance function of any two nonnegative vectors. For all vectors $x, y \geq 0$, we have for $I(x, y) \geq 0$, and the equality hold if and only if $x = y$.

Obviously, if $\beta_k = 0$, the formula (7) for the perturbed EM iteration is the scheme (6) for the classic EM iteration, in which we are not interested. Therefore, in the following we assume $\beta_k > 0$. A natural question is that under what assumptions the sequence x^n generated by the perturbed EM algorithm converges to the minimizer of (8) as well. Before discussing the convergent conditions of the perturbed EM algorithm, we first summarize some propositions of the classic EM iteration [9, 12, 24]. Without loss of generality, we assume that all the elements b_i of b are not zeroes. Furthermore, we assume $H_j = \sum_i a_{ij} = 1$ for all j in the theoretical derivation for simplicity.

Theorem 2.1 *For any initial point $x^0 > 0$, denote by x^k the estimate of the EM iteration after k iterations. Then the following propositions hold.*

1. $x_j^k > 0$ for all k, j , and $\sum_j x_j^k = \sum_i b_i, k > 0$,
2. $I_A^b(x^{k+1}) \leq I_A^b(x^k)$, and $I(x^{k+1}, x^k) \leq I_A^b(x^k) - I_A^b(x^{k+1})$,
3. $\{x^k\}$ converges to a minimizer x^* of $I_A^b(x)$ on \mathbb{R}_+^N , and x^* is a fixed point of the EM operator,
4. $I(x^*, x^{k+1}) \leq I(x^*, x^k)$.

The above inequalities hold with equalities in 2 and 4 iff x^k is a fixed point of the EM operator.

Remark 2.2 *Obviously, x is a fixed point of the EM operator iff $f_j(x) = 1$ for $x_j \neq 0$.*

From the propositions above, we have that the sequence $\{I_A^b(x^k)\}$ monotonically converges to the minimum of $I_A^b(\cdot)$ on \mathbb{R}_+^N , and $\{x^k\}$ approximates to the minimizer x^* gradually. We will investigate the convergence of the perturbed EM iteration, and prove the similar propositions of the classic EM iteration as far as we can.

Theorem 2.3 *Given any initial point x^0 , denote by $\{x^k\}_{k \in \mathbb{N}}$ the sequence generated by the perturbed EM iteration (7).*

$$x_j^{k+1} = (x_j^k + \beta_k v_j^k) \cdot f_j(x^k + \beta_k v^k) = y_j^k \cdot f_j(y^k) \quad (9)$$

where $y_j^k = x_j^k + \beta_k v_j^k$, and $\{v^k\}$ are bounded vectors. Several properties of the iterative sequence (9) can be identified.

1. $I_A^b(x^{k+1}) \leq I_A^b(x^k)$,
2. $\{x^k\}$ has a convergent subsequence $\{x^{m_k}\}$, and the limit \hat{x} is a fixed point of the EM operator,
3. furthermore, $I(\hat{x}, x^{k+1}) \leq I(\hat{x}, x^k)$, and $x^k \rightarrow \hat{x}$,
4. lastly, \hat{x} is the minimizer of $I_A^b(x)$.

The above inequalities hold with equalities iff x^k is a fixed point of the EM operator. If the following conditions hold,

1. positivity: $y_j^k > 0$,
2. decrease: $\beta_k \rightarrow 0$ and
- 3.

$$\beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} B_k^- - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} B_k^+ + \beta_k \sum_j v_j^k < I_A^b(y^k) - I_A^b(x^{k+1}), \quad (10)$$

where $S_k^- = \{j | v_j^k < 0\}$, $S_k^+ = \{j | v_j^k > 0\}$, and $B_k^- = \max\{\rho + \sum_{j \in S_k^-} x_j^{k+1}, \sum_{j \in S_k^-} \hat{x}_j\}$, $B_k^+ = \min\{\sum_{j \in S_k^+} x_j^{k+1}, \sum_{j \in S_k^+} \hat{x}_j\}$. Here ρ is a sufficiently small positive number.

Before presenting the proof, we explain the necessity of the conditions of theorem 2.3. The positive condition of y_j^k is necessary for the nonnegative constraints. The second condition is required by the convergence of iterations to a feasible point. Intuitively, the last condition is used to guarantee the decrease of the K-L distance $I_A^b(x^k)$, which means that the iterations should approximate to the minimizer of $I_A^b(\cdot)$ gradually.

An important concern is about the existence of the perturbations satisfying the conditions of theorem 2.3, especially the third condition. Because the sequence $\{\hat{x}^k\}$ generated by the classic EM iteration ($\beta_k = 0$) satisfies the conditions in theorem 2.3, the sequence $\{x^k\}$ generated by the perturbed version will also satisfy them when β_k s are small enough due to the continuities of the K-L distance (8) and the EM operator (7). Therefore, there exist perturbations which satisfy the assumptions in theorem 2.3.

Proof : We prove this theorem step by step, which follows the proving procedure in [25].

1. $I_A^b(x^{k+1}) < I_A^b(x^k)$. We will prove that the inequality (10) implies that $\{I_A^b(x^k)\}$ is a monotone decreasing sequence, i.e. $I_A^b(x^{k+1}) < I_A^b(x^k)$. By the definition of $I_A^b(\cdot)$, we have

$$\begin{aligned} & I_A^b(x^k) - I_A^b(x^{k+1}) \\ &= \sum_i b_i \log \frac{b_i}{d_i(x^k)} - \sum_i b_i \log \frac{b_i}{d_i(x^{k+1})} \\ &= \sum_i b_i \log \frac{d_i(x^{k+1})}{d_i(y^k) - \beta_k d_i(v^k)} \\ &= \sum_i b_i \log \frac{d_i(x^{k+1})}{d_i(y^k)} + \sum_i b_i \log \frac{1}{1 - \beta_k d_i(v^k)/d_i(y^k)} \\ &\geq I_A^b(y^k) - I_A^b(x^{k+1}) - \beta_k \sum_j v_j^k + \sum_i b_i \left[1 - \left(1 - \beta_k \frac{d_i(v^k)}{d_i(y^k)} \right) \right] \end{aligned} \quad (11)$$

$$\begin{aligned} &= I_A^b(y^k) - I_A^b(x^{k+1}) - \beta_k \sum_j v_j^k + \sum_i b_i \beta_k \frac{d_i(v^k)}{d_i(y^k)} \\ &\geq \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} B_k^- - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} B_k^+ + \beta_k \sum_i b_i \frac{d_i(v^k)}{d_i(y^k)} \\ &= \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} B_k^- - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} B_k^+ + \beta_k \sum_j \frac{v_j^k}{y_j^k} \sum_i \frac{a_{ij} b_i}{d_i(y^k)} \\ &= \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} B_k^- - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} B_k^+ + \beta_k \sum_j \frac{v_j^k}{y_j^k} x_j^{k+1} \\ &\geq \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} (B_k^- - \sum_{j \in S_k^-} x_j^{k+1}) - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} (B_k^+ - \sum_{j \in S_k^+} x_j^{k+1}) \\ &\geq \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} \rho - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} (B_k^+ - \sum_{j \in S_k^+} x_j^{k+1}) \end{aligned} \quad (12)$$

The first and second inequalities hold because of $\log x \geq 1 - \frac{1}{x}$ for $x > 0$ and the condition (10), respectively. We have proved that $I_A^b(x^{k+1}) \leq I_A^b(x^k)$ by (12) and condition 3. Next we should prove that $I_A^b(x^{k+1}) = I_A^b(x^k)$ iff $\beta_k v^k = 0$ and x^k is a fixed point further. Obviously, the sufficiency is true by theorem 2.1. The necessity is also true by the following facts. If $I_A^b(x^{k+1}) = I_A^b(x^k)$, we have that all the inequalities in derivation above hold with equalities, which imply that $\frac{1}{1 - \beta_k d_i(v^k)/d_i(y^k)} = 1$ and $S_k^- = \emptyset$ by inequalities (11) and (12), respectively, i.e. $\beta_k d_i(v^k) = 0$ and $v_j^k \geq 0$. In fact, we have $v_j^k = 0 (\forall j)$, i.e. $y_j^k = x_j^k$, and that x^k is a fixed point by theorem 2.1. If there exists j_0 such that $v_{j_0}^k > 0$, there must exist i_0 such that $a_{i_0 j_0} > 0$ and $d_{i_0}(v^k) = \sum_j a_{i_0 j} v_j^k \geq a_{i_0 j_0} v_{j_0}^k > 0$ by the assumption $\sum_i a_{ij} = 1 \neq 0$, which is contradict to $d_i(v^k) = 0 (\forall i)$.

2. **Existing convergent subsequence $\{x^{m_k}\}$:** Because $0 < \sum_j x_j^k = \sum_i b_i$ (constant) and $x_j^k > 0$, $\{x^k\}$ has a convergent subsequence $\{x^{m_k}\}$. Denoting by \hat{x} the limit of $\{x^{m_k}\}$, we have $y^{m_k} \rightarrow \hat{x}$ because $\beta_k \rightarrow 0$ and v^k is bounded.

Next we will prove \hat{x} is a fixed point of EM operator. To this end, we define a function about $x \geq 0$

$$D(x) = I(P(x), x). \quad (13)$$

By the second proposition of theorem 2.1, we have

$$D(y^{m_k}) = I(P(y^{m_k}), y^{m_k}) = I(x^{m_k+1}, y^{m_k}) \leq I_A^b(y^{m_k}) - I_A^b(x^{m_k+1}). \quad (14)$$

Due to the statement 1, $\{I_A^b(x^k)\}$ is a convergent sequence. Furthermore, $I_A^b(y^k)$ converges to the same limit as $I_A^b(x^k)$ because of the continuity of $I_A^b(\cdot)$ and $\beta_k \rightarrow 0$. Therefore, we have $I_A^b(y^{m_k}) - I_A^b(x^{m_k+1}) \rightarrow 0$, $m_k \rightarrow \infty$. Thus, we have that

$$D(\hat{x}) = \lim_{m_k \rightarrow \infty} D(y^{m_k}) = I(P(\hat{x}), \hat{x}) = 0, \quad (15)$$

i.e. \hat{x} is a fixed point of the EM operator, since $I(x, y) = 0 \iff x = y$. Therefore, we have $f_j(\hat{x}) = 1$ if $\hat{x}_j \neq 0$ by the EM iteration formula.

3. $I(\hat{x}, x^{k+1}) \leq I(\hat{x}, x^k)$. Before going further, we first prove a general inequality $I(\hat{x}, x) - I(\hat{x}, P(x)) \geq I_A^b(x) - I_A^b(\hat{x})$ for any point $x > 0$. Assume $z = P(x)$, then we have

$$\begin{aligned} & I(\hat{x}, x) - I(\hat{x}, z) \\ &= \sum_j \hat{x}_j \log \frac{z_j}{x_j} - \sum_j (\hat{x}_j - x_j) \\ &= \sum_j \hat{x}_j \sum_i g_{ij}(\hat{x}) \log \frac{z_j}{x_j} \frac{g_{ij}(x)}{g_{ij}(\hat{x})} - \sum_j (\hat{x}_j - x_j) \\ &= \sum_j \hat{x}_j \sum_i g_{ij}(\hat{x}) \log \frac{z_j}{x_j} \frac{g_{ij}(\hat{x})}{g_{ij}(x)} + \sum_j \hat{x}_j \sum_i g_{ij}(\hat{x}) \log \frac{g_{ij}(x)}{g_{ij}(\hat{x})} - \sum_j (\hat{x}_j - x_j) \\ &\geq - \sum_j \hat{x}_j \log \sum_i g_{ij}(\hat{x}) \frac{x_j}{z_j} \frac{g_{ij}(x)}{g_{ij}(\hat{x})} + \sum_j \hat{x}_j \sum_i g_{ij}(\hat{x}) \log \frac{\frac{a_{ij} b_i}{d_i(x)}}{\frac{a_{ij} b_i}{d_i(\hat{x})}} - \sum_j (\hat{x}_j - x_j) \\ &= - \sum_j \hat{x}_j \log \frac{x_j f_j(x)}{z_j} + \sum_i \sum_j \hat{x}_j \frac{a_{ij} b_i}{d_i(\hat{x})} \log \frac{d_i(\hat{x})}{d_i(x)} - \sum_j (\hat{x}_j - x_j) \\ &= \sum_i b_i \log \frac{d_i(\hat{x})}{d_i(x)} - \sum_j (\hat{x}_j - x_j) \\ &= \sum_i b_i \ln \frac{b_i}{d_i(x)} - \sum_i b_i \ln \frac{b_i}{d_i(\hat{x})} - \sum_i (b_i - d_i(x)) + \sum_i (b_i - d_i(x)) - \sum_j (\hat{x}_j - x_j) \\ &= I_A^b(x) - I_A^b(\hat{x}), \end{aligned} \quad (16)$$

where we used $\sum_j \hat{x}_j = \sum_j x_j^k = \sum_i b_i = \sum_j z_j$ and the definition of $I(\cdot, \cdot)$. The inequality and the second equality were implied by Jensen's inequality and the fact $f_j(\hat{x}) = \sum_i g_{ij}(\hat{x}) = 1$ for $\hat{x}_j \neq 0$, respectively. The last equality follows $\sum_j \hat{x}_j = \sum_i b_i$ and $\sum_i d_i(x) = \sum_i \sum_j a_{ij} x_j = \sum_j x_j$ by the assumption $\sum_i a_{ij} = 1$ and the property 1 of theorem 2.1. Let $x = y^k$, then $z = P(y^k) = x^{k+1}$ and (16)

implies

$$\begin{aligned}
I(\hat{x}, y^k) - I(\hat{x}, x^{k+1}) &\geq I_A^b(y^k) - I_A^b(\hat{x}) \\
&= [I_A^b(y^k) - I_A^b(x^{k+1})] + I_A^b(x^{k+1}) - I_A^b(\hat{x}) \\
&\geq I_A^b(y^k) - I_A^b(x^{k+1}).
\end{aligned} \tag{17}$$

The last inequality held because $I_A^b(x^{k+1}) \geq I_A^b(\hat{x})$.

Next we have prove the property of $\{x^k\}$ that $I(\hat{x}, x^{k+1}) \leq I(\hat{x}, x^k)$. By the definitions of the K-L distance and the perturbed EM formula, we have

$$\begin{aligned}
&I(\hat{x}, x^k) - I(\hat{x}, x^{k+1}) \\
&= \sum_j \hat{x}_j \log \frac{(x_j^k + \beta_k v_j^k) f_j(y^k)}{x_j^k} \\
&= \sum_j \hat{x}_j \log f_j(y^k) + \sum_j \hat{x}_j \log(1 + \beta_k \frac{v_j^k}{x_j^k}) \\
&\geq \sum_j \hat{x}_j \log \frac{y_j^k f_j(y^k)}{y_j^k} + \sum_j \hat{x}_j \left(1 - \frac{1}{1 + \beta_k v_j^k / x_j^k}\right) \\
&= I(\hat{x}, y^k) - I(\hat{x}, x^{k+1}) + \sum_j \hat{x}_j \frac{\beta_k v_j^k}{x_j^k + \beta_k v_j^k} - \beta_k \sum_j v_j^k
\end{aligned} \tag{18}$$

$$\begin{aligned}
&\geq I_A^b(y^k) - I_A^b(x^{k+1}) - \beta_k \max_{j \in S_k^-} \{-\frac{v_j^k}{y_j^k}\} \sum_{j \in S_k^-} \hat{x}_j + \beta_k \min_{j \in S_k^+} \{\frac{v_j^k}{y_j^k}\} \sum_{j \in S_k^+} \hat{x}_j - \beta_k \sum_j v_j^k \\
&\geq \beta_k \max_{j \in S_k^-} \{-\frac{v_j^k}{y_j^k}\} (B_k^- - \sum_{j \in S_k^-} \hat{x}_j) + \beta_k \min_{j \in S_k^+} \{\frac{v_j^k}{y_j^k}\} (\sum_{j \in S_k^+} \hat{x}_j - B_k^+) \\
&\geq 0,
\end{aligned} \tag{19}$$

Again we used the inequality $\log x \geq 1 - \frac{1}{x}$ for $x > 0$ for the first inequality. The second and the third inequalities hold by (17) and amplification and shrink technique. The last inequality hold by the formula (10).

We have that $I(\hat{x}, x^{k+1}) \leq I(\hat{x}, y^k)$ by equation (19). Furthermore, we have $I(\hat{x}, x^{k+1}) = I(\hat{x}, y^k)$ iff $v_j^k = 0$ and x^k is the fixed point of the EM operator. The sufficiency is clear.

The necessity is also true by the following facts. If $I(\hat{x}, x^{k+1}) = I(\hat{x}, y^k)$, we have that all the inequalities in deviation above hold with equalities, which imply that $v_j^k = 0$ and $I_A^b(y^k) = I_A^b(x^{k+1})$ by inequalities (18) and (19), respectively. Therefore $y^k = x^k$ and x^k is the fixed point of the EM operator by the theorem 2.1. Now we prove the convergence of $\{x^k\}$ generated by the perturbed EM formula, i.e $x^k \rightarrow \hat{x}$. From (19), we can conclude that $I(\hat{x}, x^k)$ is a monotone decreasing sequence if the assumptions of theorem 2.3 hold. Thus, we have $I(\hat{x}, x^k) \searrow 0$ because of $I(\hat{x}, x^{m_k}) \rightarrow 0$ and the monotonicity of $I(\hat{x}, x^k)$. Therefore, we have proved $x^k \rightarrow \hat{x}$, and $y^k \rightarrow \hat{x}$ by $\beta_k v^k \rightarrow 0$ and $I(x, y) = 0 \implies x = y$.

4. **\hat{x} is a minimizer of $I_A^b(\cdot)$ on \mathbb{R}_+^n .** In order to prove this statement, we should prove \hat{x} satisfies the Kuhn-Tucker (K-T) conditions. For the EM algorithm and the K-L function, the K-T conditions are equivalent to (1) $f_j(\hat{x}) = 1$ if $\hat{x}_j \neq 0$ and (2) $0 \leq f_j(\hat{x}) \leq 1$ if $\hat{x}_j = 0$ [9, 25]. Obviously, the firstly condition holds because \hat{x} is the fixed point of the EM operator. For the second condition, the nonnegativity of \hat{x} is satisfied since \hat{x} is the limit of $x^k \geq 0$. Next we prove that $f_j(\hat{x}) \leq 1$ for $\hat{x}_j = 0$. If $f_j(\hat{x}) \geq 1 + \epsilon > 1$ for $\hat{x}_j = 0$, by the continuities of $f_j(\cdot)$ and the EM operator and $y^k \rightarrow \hat{x}$, we have that there exists a sufficiently large integer $L > 0$ such that $f_j(y^k) > 1 + \epsilon_1$ with $\epsilon_1 > 0$ for all $k \geq L$. If $x_j^k \rightarrow 0$, there must be infinite iterations such that $x_j^{k+1} < x_j^k$. In the following, we assume that $k > L$. For each iteration satisfying $x_j^{k+1} = y_j^k \cdot f_j(y^k) < x_j^k$, we have that

$$\begin{aligned}
1 &> \frac{(x_j^k + \beta_k v_j^k) f_j(y^k)}{x_j^k} \\
&= (1 + \beta_k \frac{v_j^k}{x_j^k}) f_j(y^k)
\end{aligned} \tag{20}$$

By the assumption $f_j(y^k) > (1 + \epsilon_1)$, we have

$$1 + \beta_k \frac{v_j^k}{x_j^k} < \frac{1}{1 + \epsilon_1}. \quad (21)$$

Abstracting 1 on both sides of (21), we have $\beta_k \frac{v_j^k}{x_j^k} < -\frac{\epsilon_1}{1 + \epsilon_1}$. By equation (12), we have that

$$\begin{aligned} & I_A^b(x^k) - I_A^b(x^{k+1}) \\ & \geq \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} (B_k^- - \sum_{j \in S_k^-} x_j^{k+1}) - \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} (B_k^+ - \sum_{j \in S_k^+} x_j^{k+1}) \\ & \geq \frac{\epsilon_1}{1 + \epsilon_1} (B_k^- - \sum_{j \in S_k^-} x_j^{k+1}) \\ & \geq \rho \frac{\epsilon_1}{1 + \epsilon_1} \end{aligned} \quad (22)$$

The second inequality hold because $y_j^k = x_j^k + \beta_k v_j^k < x_j^k < 0$ and $B_k^+ \leq \sum_{j \in S_k^+} x_j^{k+1}$. The last inequality follows from the definition of B_k^- .

From equation (22), we can see that the decrease of $I_A^b(x^k)$ is larger than a positive number for each iteration satisfying $x_j^{k+1} < x_j^k$. Further, because the number of iterations satisfying $x_j^{k+1} < x_j^k$ is infinite and $I_A^b(x^k)$ is finite number, $I_A^b(x^l)$ will go to negative infinity from, which is contradict to the fact that $I_A^b(x^l) \geq 0$ for all x^l .

3 Superiorization of the EM Algorithm

The regularization reconstruction for SPECT is often modeled as a constrained optimization problem

$$\min_{x \in E} \phi(x), \quad E = \{x^* | x^* = \arg \min_{x \geq 0} I_A^b(x)\}, \quad (23)$$

where ϕ is a convex function, which assigns an image x a number indicating the "undesirability" of the image in some sense. The set E is called feasible set, and it is called feasible problem to find a point in E .

To our knowledge, there is no efficient algorithm to deal with the constrained optimization problem (23) because of the large scale of it for SPECT reconstruction. On the other hand, although the feasible problem is also a constrained optimization, we can solve it by the classic EM algorithm [9] and its variants [10, 11] efficiently. Based on the facts above, we use the superiorization methodology for the given objective function to implement the regularization reconstruction of SPECT.

For the objective function ϕ , the superiorized EM algorithm is illustrated as algorithm 1 based on the conditions of theorem 2.3. In order to emphasize the objective function ϕ for which we are superiorizing, we refer to the superiorized algorithm as the ϕ -superiorization of the EM iteration.

Algorithm 1 Framework of ϕ -superiorization algorithm

```

Initialization:  $\beta_0 > 0$ ,  $x^0 > 0$ ,  $k = 0$ , and  $0 < \gamma < 1$ .
repeat: logic=true
  while logic
    find a decreasing direction  $v^k$  of  $\phi$  at  $x^k$ , such that  $y_j^k = v_j^k + \beta_k v_j^k > 0$ ;
    If  $\phi(y^k) \leq \phi(x^k)$  and inequality (10) hold (*)
      logic=false,  $x^{k+1} = P(y^k)$ ,  $\beta_{k+1} = \beta_k$ ,  $k = k + 1$ 
    end(if)
     $\beta_k = \gamma \beta_k$ 
  end(while)

```

Because the perturbation direction v^k is selected as the decreasing direction of ϕ at x^k , the size of β_k represents the strength of regularization in some sense. Because the condition 3 of theorem 2.3 is very strict, the numerical experiments show that $\{\beta_k\}$ goes to zero very fast, which results the the regularization effect is very weak. Therefore, we propose a modified version of algorithm 1, which is shown in algorithm 2. In algorithm 2, we only validate $I_A^b(x^{k+1}) < I_A^b(x^k)$, rather than the inequality (10) of theorem 2.3. Since the inequality (10) implies $I_A^b(x^{k+1}) < I_A^b(x^k)$ by theorem 2.3, algorithm 2 can be seen as a relaxation of algorithm 1. Furthermore,

we introduce a relative decrease of $I_A^b(\cdot)$ to avoid the situation that the amount of $I_A^b(x^k) - I_A^b(x^{k+1})$ for each iteration is too small, which can accelerate the convergence of $I_A^b(x^k)$ and x^k intuitively.

In order to confirm the inequality (10) of theorem 2.3 in algorithm 1, we should compute $\sum_{j \in S_k^-} \hat{x}_j$ and

Algorithm 2 Modification of the ϕ -superiorization algorithm 1

```

 $\beta_0 > 0, x^0 > 0, k = 0$ , and  $0 < \gamma < 1$ .
repeat: logic=true
  while logic
    find a decreasing direction  $v^k$  of  $\phi$  at  $x^k$ , such that  $y_j^k = v_j^k + \beta_k v_j^k > 0$ ;
    If  $\phi(y^k) \leq \phi(x^k)$  and  $I_A^b(Py^k) < I_A^b(x^k)$  (★)
      logic=false,  $x^{k+1} = P(y^k)$ .
      If  $\frac{I_A^b(x^k) - I_A^b(x^{k+1})}{I_A^b(x^k)} < H_1$  (a1)
         $\beta_{k+1} = \gamma \beta_k$  (a2)
      else
         $\beta_{k+1} = \beta_k$ 
      end(else)
       $k = k + 1$ 
    else
       $\beta_k = \gamma \beta_k$  (a3)
    end(else)
  end(while)

```

$\sum_{j \in S_k^+} \hat{x}_j$ for each iteration. However, the limit \hat{x} is not known in the computational procedure. In practice, we estimate the values $\sum_{j \in S_k^-} \hat{x}_j$ and $\sum_{j \in S_k^+} \hat{x}_j$ by

$$\sum_{j \in S_k^-} \hat{x}_j \approx \frac{|S_k^-|}{N} B, \quad (24)$$

$$\sum_{j \in S_k^+} \hat{x}_j \approx \frac{|S_k^+|}{N} B, \quad (25)$$

where $|S_k^-|$ and $|S_k^+|$ denote the cardinalities of the sets S_k^- and S_k^+ , respectively. Furthermore, the capital letters N and $B = \sum_i b_i$ in the formulae (24) and (25) denote the length of x and the total counts, respectively. Further, the parameter ρ in theorem 2.3 is chosen as 0.01 and 10^{-6} .

In the following, we will discuss how to generate desirable perturbation $\beta_k v^k$ or y^k for two concrete objective functions, TV and l^1 -norm, such that the perturbations satisfy the conditions (*) and (★) of the two ϕ -superiorization algorithms. Since the TV regularization allows the reconstructed image to have sharp edges, TV based models are widely used in imaging sciences [15, 22, 26, 27]. For an $H \times W$ image x whose pixel values are denoted by $x_{i,j}$, the TV of x is defined as

$$TV(x) = \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}. \quad (26)$$

where H, W are the height and width of x . In order to decrease the value of TV at x^k , we choose v^k as

$$v^k = s^k / |s^k| \quad (27)$$

where $s^k \in \partial TV(x^k)$ is the sub-gradient of TV at point x^k , and $|s^k|$ is the maximum absolute value of the components of s^k . Therefore, the sequence $\{v^k\}$ is bounded. In fact, v^k is the normalization of s^k in the l^∞ space, rather than in the l^2 space used in [18, 20].

In addition to the TV based models, l^1 -norm minimization method is another widely used technique in image sciences and processing [19][28][29]. Here the l^1 -norm is about the wavelet coefficients of x , which is defined as

$$\|T_{\{\psi_j\}}(x)\|_1 = \sum_j |\alpha_j| \quad (28)$$

where α_j s are the coefficients of x under a given wavelet basis $\{\psi_j\}$, and the letter T denotes the wavelet decomposition operator. Although we can use the same method for TV function to reduce the wavelet l^1 -norm

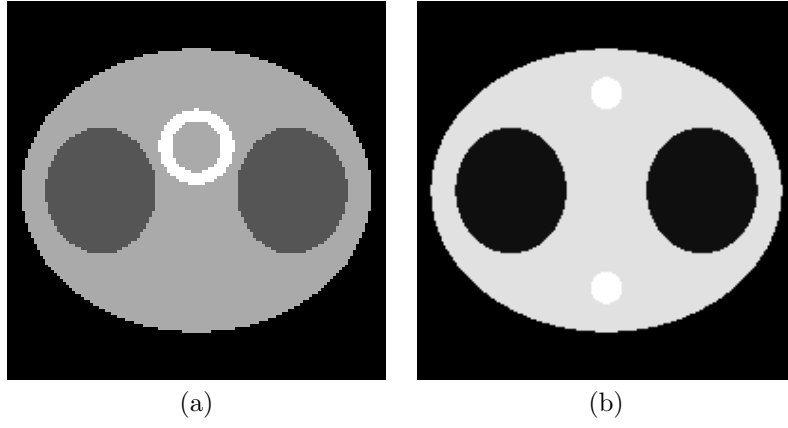


Figure 1: (a): activity map (b): attenuation map.

of x^k , we introduce two more effective methods, soft and hard thresholding schemes, to decrease the wavelet l^1 -norm of x^k . Define γ_j^k by

$$\gamma_j^k = \text{Hard}(\alpha_j^k) = \begin{cases} 0 & |\alpha_j^k| \geq \beta_k \\ -\frac{\alpha_j^k}{\beta_k} & |\alpha_j^k| < \beta_k \end{cases}, \quad (29)$$

$$\gamma_j^k = \text{Soft}(\alpha_j^k) = \begin{cases} -\text{sign}(\alpha_j^k) & |\alpha_j^k| \geq \beta_k \\ -\frac{\alpha_j^k}{\beta_k} & |\alpha_j^k| < \beta_k \end{cases}, \quad (30)$$

where α_j^k are the wavelet coefficients of x^k . The perturbation direction for each iteration is selected as $v^k = T_{\{\psi_j\}}^{-1}(\gamma_j^k)$, and $y^k = x^k + \beta_k v^k$. In fact, we need not to compute v^k explicitly. By the linearity of wavelet transform, we can compute the wavelet coefficients of y^k directly by

$$\alpha_j^k = \text{Hard}^C(\alpha_j^k) = \begin{cases} \alpha_j^k & |\alpha_j^k| \geq \beta_k \\ 0 & |\alpha_j^k| < \beta_k \end{cases}, \quad (31)$$

$$\alpha_j^k = \text{Soft}^C(\alpha_j^k) = \begin{cases} \alpha_j^k - \text{sign}(\alpha_j^k) \cdot \beta_k & |\alpha_j^k| \geq \beta_k \\ 0 & |\alpha_j^k| < \beta_k \end{cases}, \quad (32)$$

and $y^k = T_{\{\psi_j\}}^{-1}(\alpha_j^k)$. In the numerical simulations, we use Daubechies 6.8 biorthogonal wavelets with symmetric extensions at the boundaries. For refereed convenience, we use hard-superiorization and soft-superiorization to distinguish them for l^1 -superiorization in this paper. In order to avoid $y_j^k \leq 0$, y_j^k is set as $\frac{1}{2}x_j^k$ if $y_j^k \leq 0$ in practice.

4 Numerical Results

In this section we investigate the stability and robust of the proposed algorithms by several numerical experiments of SPECT reconstruction. To this end, we conducted simulation studies by employing computer, and the projection data were generated based on the following model. Figure 1 (a) and (b) display a chest phantom of activity and an attenuation map, respectively. The chest phantom activity consists of an ellipsoidal background (body region) with axes of length 22.5cm and 30cm, which contains two smaller ellipsoidal regions(lungs) with axes of length 10cm and 8.8cm, and a ring(myocardium) which inner and outer diameters are 6cm and 8cm, respectively. The activity in myocardium, background, and lungs were specified to be in the ratio 8:3:1.

To simulate the strongly non-uniform attenuation coefficient of chest we utilize the phantom used in [10], which imitates the attenuation map across a section of human thorax. Besides the body region and the lungs of the activity map, the attenuation map consists of two circular regions (bones) of diameter 2.5cm(see figure 1(b)). Attenuation coefficients were 0.03cm^{-1} within 'lung' regions, 0.17cm^{-1} within 'bone' regions, 0.15cm^{-1} elsewhere within the body ellipse, and 0.00cm^{-1} outside the body.

In the simulation, the activity and attenuation maps are evenly sampled in $[-15, 15] \times [-15, 15]$ on a grid of 128×128 . A perfect parallel hole collimator was assumed, and noise-free projection data were created via attenuated Radon transform formula (1), which included tissue attenuation, but scattering and blurring. In order to simulate the quantum noise in the simulated data, the following procedure was implemented for each data set [31].

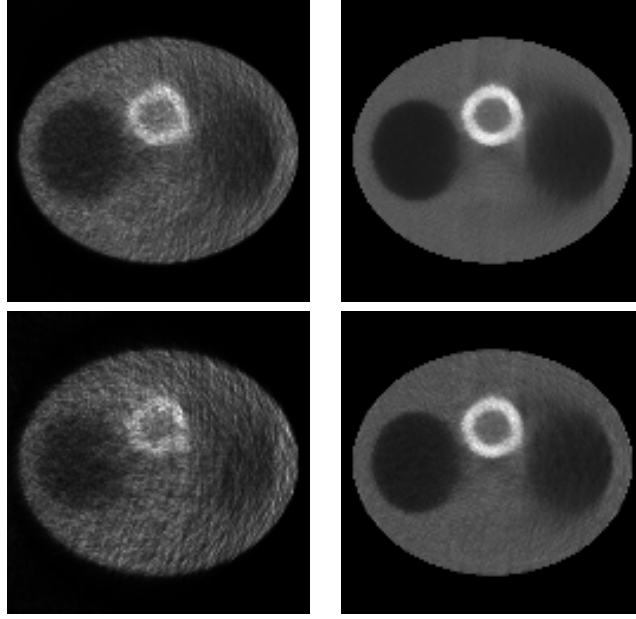


Figure 2: Images reconstructed from the simulated data set 1(top row) and 2(bottom) by the classic EM algorithm(left column) and the mean images of 100 trials of noise(right column).

- The projections were scaled (multiplied by a constant factor) so that the total count was a predefined integer,
- each point value in the data set was then replaced by a random realization of a Poisson variate with a mean equal to that value.

Two data sets, sixty and thirty projections, were generated over 180° evenly with view angles $\varphi_l = \frac{l-1}{N_0}\pi$ ($l = 1, \dots, N_0$, and $N_0 = 30$ or 60). Counts were recorded in 128 bins per projection. Total counts recorded on all projections are about 500K and 100K for two simulation data sets.

In order to evaluate the performance of the superiorized algorithms, we computed the mean image x^* from 100 noise trials of the two data sets.

$$x^* = \frac{1}{100} \sum_{m=1}^{100} \hat{x}^m, \quad (33)$$

where \hat{x}^m ($i = 1, 2, \dots, 100$) is the reconstructed image of the m th noise trial by the classic EM algorithm, and the number of iteration for each trial is 30.

Figure 2 displays the reconstructed images by the classic EM algorithm and the mean images for two simulated data sets. As expected, the mean images for the two data sets are clear visually comparing with the images reconstructed by the classic EM algorithm, which are dominated by noise, especially the image for data set 2 due to the lower counts. Thus, we used the mean square error(MSE) between the estimations and the mean images for two data sets to measure the image qualities. And we used it as the criteria to select the outputs of the iterative algorithms. The mean square error(MSE) between the estimation x and the mean estimation is computed by

$$\text{MSE}(x) = \frac{1}{N} \sum_{j=1}^N (x_j - x_j^*)^2 \quad (34)$$

The outputs of the different algorithms are taken as the best estimations in terms of MSE. In order to compare the image quality from different data set, we introduce an amount of relative MSE(RMSE), which is defined as

$$\text{RMSE}(x) = \sqrt{\frac{\sum_{j=1}^N (x_j - x_j^*)^2}{\sum_{j=1}^N (x_j^*)^2}} \quad (35)$$

where x^* is the mean image of the same data set.

In the numerical simulations, an initial image x^0 with value $c = \frac{1}{N} \sum_i b_i$ is used for all experiments, unless

there is a further explanation. The parameters β_0 are chosen as $c/2$ and $c/10$ for the TV- and l^1 -superiorized EM algorithms, respectively. The threshold H_1 of the relative decrease of $I_A^b(x^k)$ in the superiorized algorithm 2 and the parameter γ are chosen as 0.01 and $1/2$, respectively.

Experiment 1: sixty projections and 500K counts

Figure 3 displays the images reconstructed from data set 1 by the TV-, hard- and soft-superiorized EM algorithms from column 1 to column 3, respectively. Figure 4 plots the variances of $\text{MSE}(x^k)$, $I_A^b(x^k)$ and β_k versus the iteration number. Table 1 tabulates the TV values, l^1 -norms, RMSEs and iterations of the images in figure 3, the image by the classic EM algorithm and the mean image for data set 1 in figure 2. In the following, TV-, hard- and soft- n ($n = 1, 2$) are the abbreviations of the TV-, hard and soft-superiorized EM algorithm n ($n = 1, 2$) for simplicity.

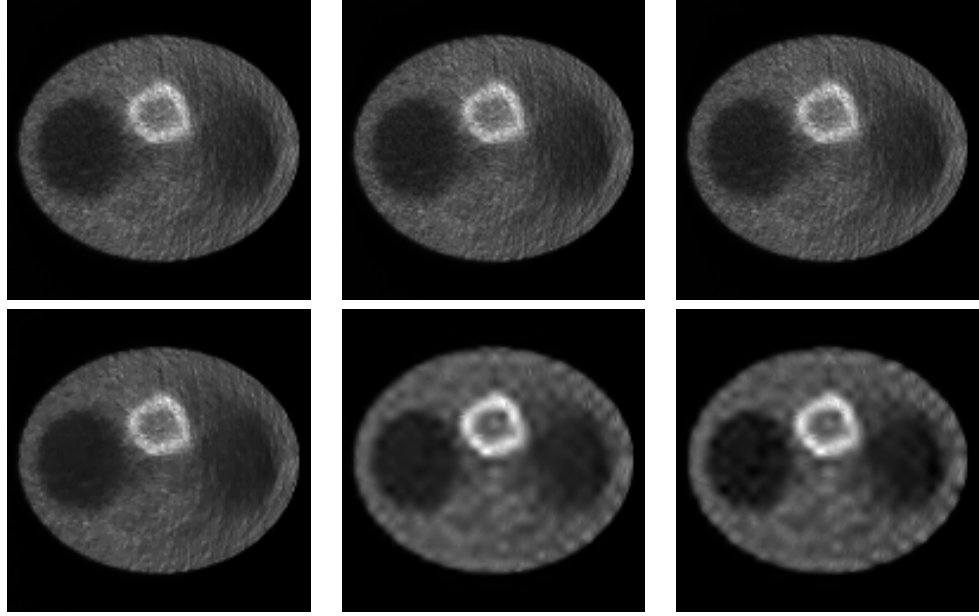


Figure 3: Images reconstructed from data set 1. The images in top and bottom rows are reconstructed by the superiorized EM algorithm 1 and algorithm 2, respectively. The images reconstructed by TV-, hard- and soft-superiorized EM algorithms are displayed from column 1 to column 3.

Table 1: TV, l^1 -norm, RMSE and iteration of the images in figure 3.

image	EM	TV-alg1	hard-alg1	soft-alg1
TV($\times 10^3$)	26.708	25.318	25.577	25.958
$l^1(\times 10^3)$	12.953	12.404	12.320	12.489
RMSE	0.1914	0.1873	0.1875	0.1888
iterations	13	13	13	13
image	mean	TV-alg2	hard-alg2	soft-alg2
TV($\times 10^3$)	13.987	19.434	14.115	15.443
$l^1(\times 10^3)$	8.5244	10.279	2.867	3.189
RMSE	-	0.1633	0.1563	0.1747
iterations	-	15	30	23

As expectation, we can see that the images by the superiorized EM algorithms are superior to the one by the classic EM algorithm from the comparison of TV, l^1 -norm and RMSE values in table 1, though the images by the superiorized algorithm 1 are visually indistinguishable from the one by the classic EM algorithm. Furthermore, the ϕ -superiorized EM algorithm 2 are superior to the corresponding ϕ -superiorized EM algorithm 1. And we can see that the results by the l^1 -superiorized EM algorithms 2 are superior to the TV-superiorized algorithm 2 visually. The reason for above phenomena maybe that the parameter β_k goes to zero very fast for the superiorized EM algorithm 1 and the TV-superiorized EM algorithms 2 (see figure 4).

From figure 4, we can see that the $I_A^b(x^k)$ decreased if β_k and v^k satisfied the conditions 3 of theorem 2.3, which validate the conclusions of the theorem 2.3 numerically. Further, we can see the plots of $I_A^b(x^k)$ for the classic EM algorithm and the superiorized EM algorithms 1 are indistinguishable from each other because

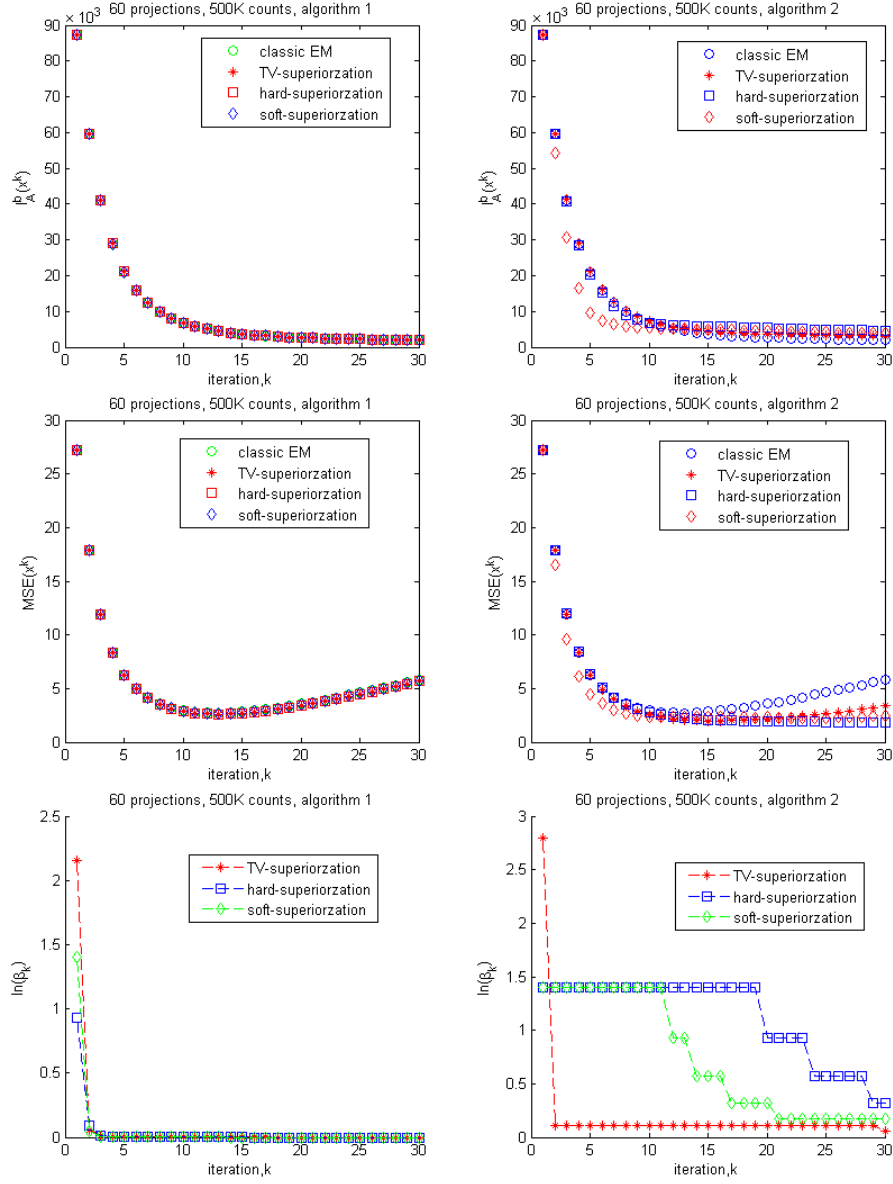


Figure 4: Plots of the K-L distance $I_A^b(x^k)$ (top row), mean square error $MSE(x^k)$ (middle row) and β_k (bottom row) versus the iterations k of the classic EM iteration and the superiorized versions(left for algorithm 1 and right for algorithm 2).

the parameters β_k went to zero very fast for the superiorized algorithm 1. The reason for the fast decreasing of β_k for the superiorized algorithm 1 is that the inequality for condition 3 is very strict because of the use of magnification technique several times. Lastly, we can see that the plots of $MSE(x^k)$ for the classic EM algorithm and the superiorized version 1 are presented as L -curve: decrease first and then increase, though $I_A^b(x^k)$ decreased gradually. And we can see that the plots of $MSE(x^k)$ for the l^1 -superiorized EM algorithm 2 decreased unceasingly, which implies the estimate x^k for each iteration approximate the mean image gradually.

From table 1, we can see that the iterations of the superiorized EM algorithm 2 is larger than the classic EM and the superiorized EM algorithm 1. However, the RMSE of the estimation of the TV-, hard- and soft-superiorized algorithms 2 after 13 iterations are 0.1668 0.1741 0.1797, respectively. Therefore, the superiorized algorithms 2 are superior to the classic EM and the superiorized EM algorithms.

Experiment 2: thirty projections and 100K counts.

Figure 5 displays the images reconstructed from data set 2 by the TV-, hard- and soft-superiorized EM algorithms 1 and 2 from column 1 to column 3 and from row 1 to row 2. Figure 6 plots the variances of $MSE(x^k)$, $I_A^b(x^k)$ and β_k versus the iteration number. Table 2 tabulates the TV values, l^1 -norms, RMSEs and iterations of the images in figure 5, the image by the classic EM algorithm and the mean image for data set 2(see figure 2).

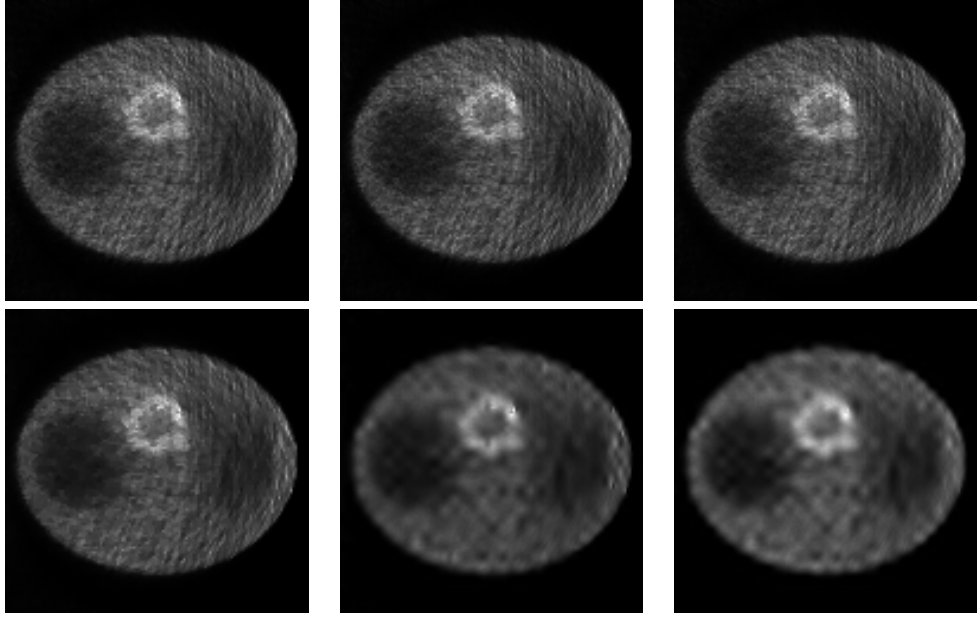


Figure 5: Images reconstructed from data set 2. The images reconstructed by the TV-, hard- and soft-superiorized EM algorithm 1 and 2 are displayed in column 2, 3 and 4, and row 1 and 2, respectively.

Table 2: TV, l^1 -norm, RMSE and iteration of the images in figure 5.

image	EM	TV-alg1	hard-alg1	soft-alg1
TV($\times 10^3$)	13.739	13.105	13.396	13.061
$l^1(\times 10^3)$	6.729	6.452	6.451	6.277
RMSE	0.2822	0.2778	0.2802	0.2795
iterations	8	8	8	8
image	mean	TV-alg2	hard-alg2	soft-alg2
TV($\times 10^3$)	7.167	9.769	5.733	5.576
$l^1(\times 10^3)$	4.037	5.069	1.434	1.283
RMSE	0	0.2490	0.2091	0.2064
iterations	-	9	12	10

From the observations of figure 5 and 6, and table 2, we can draw the same conclusions as the experiment 1. Further, we can see that the visual quality are inferior to the images of experiment 1 by comparing with figure 3 because of the low counts level. And the superiority of the ϕ -superiorized EM algorithm 2, especially the l^1 -superiorized algorithm 2, is much more remarkable than the experiment 1.

For the two experiments above, though the l^1 -superiorized EM algorithms obtain better results than the classic EM algorithm and the TV-superiorized versions, the reconstructed images by l^1 -superiorized EM algorithms contains Gibbs phenomenon caused by the hard- and soft-threshold techniques.

From figure 2, 3 and 5, we can see that the right part of the images are blurred strongly. This is because the detectors rotated from up to bottom through right, and the the gamma rays emitted from the right hand pixels were more likely to be absorbed in propagation.

In the following experiment 3, we will see the outstanding performance of the superiorized algorithm 2 for initial vector with random elements.

Experiment 3: initial x^0 with random values on interval $[1, 2]$ for data set 1

This experiment is used to test the stability and robustness of the superiorized EM algorithms for initial point x^0 . In this experiment, the initial point x^0 is selected as a random vector with values on interval $[1, 2]$. The reconstructed images by different algorithms are displayed in figure 7, and the corresponding TV values, l^1 -norms and RMSEs are tabulated in table 3.

We observe that the TV- and l^1 -superiorized EM algorithms 2 are stable and robust for initial point. By comparing the figure 3 and 7, and table 1 and 3, we have that the outputs of the classic EM algorithm and the superiorized algorithm 1 are strongly effected by the initial points, while the effect of initial point for the

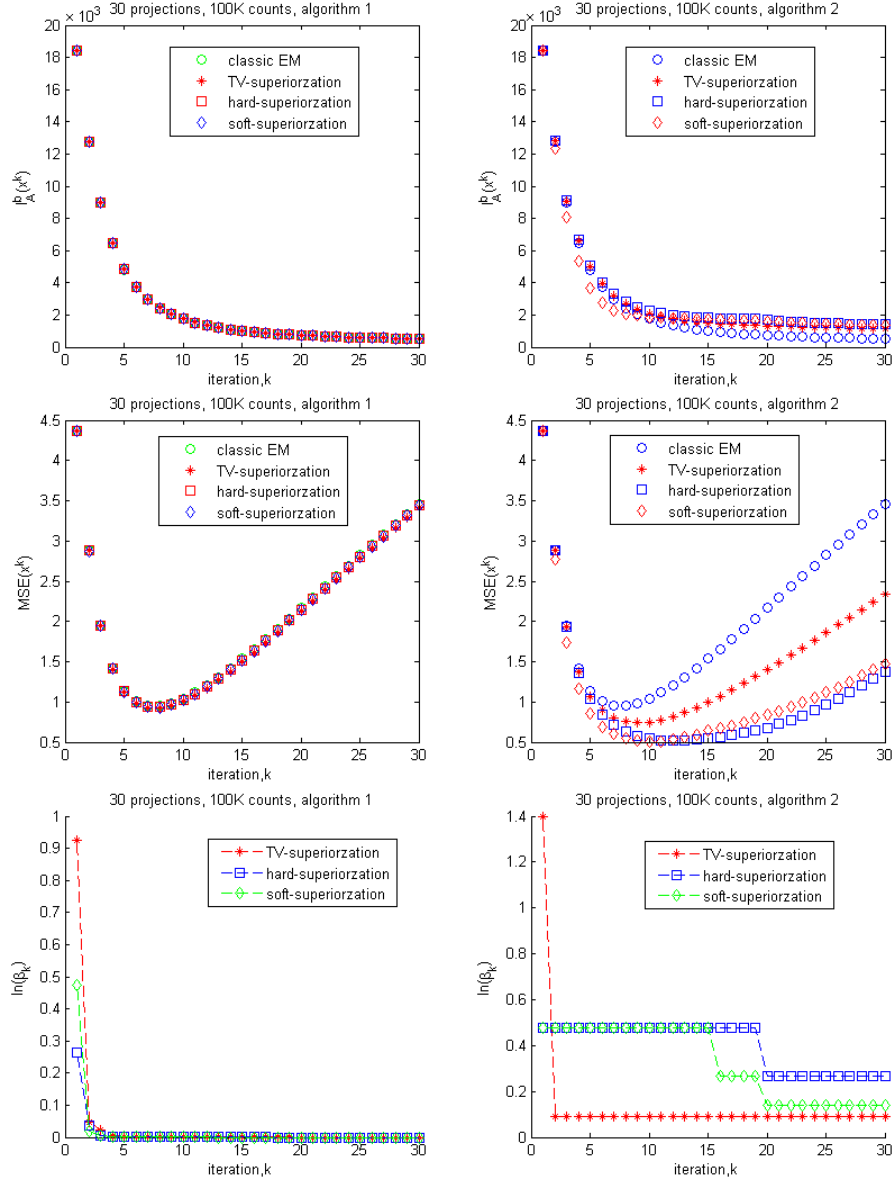


Figure 6: The K-L distance $I_A^b(x^k)$ (top row), mean square error $MSE(x^k)$ (middle row) and β_k (bottom row) variance versus the iterations k of the classic EM iteration and the superiorized versions(left for algorithm 1 and right for algorithm 2) for experiment.

TV- and l^1 -superiorized EM algorithm 2 is very weak. Because the variances of $I_A^b(x^k)$ and $MSE(x^k)$ versus iteration are very similar to that of experiment 1, we only plot the variance of β_k in figure 8.

A surprising observation is that the random initial vector is superior to the uniformly initial vector for the superiorized EM algorithms 2 in term of RMSE by comparing the table 1 and 3. This is contrary to the long-standing initial selection of iterative method for image reconstruction.

By comparing the plots of β_k of the experiments above for different algorithms, we can observe the following facts. Firstly, the parameter β_k went to zero very fast for the superiorized EM algorithms 1. Secondly, the parameters β_k of experiment 1 and 2 went to zero slowly for the l^1 -superiorized algorithm 2 while fast for the TV-superiorized EM algorithm 2 with uniformly initial point(see figure 4 and 6). Thirdly, the parameter β_k of experiment 3 went to zero at a relatively slow rate for the TV-superiorized EM algorithm 2 comparing with experiment 1 and 2.

The reasons for this include the following aspects. Firstly, as discussed above, the parameter β_k for the superiorized EM algorithm 1 went to zero fast by the strict condition (3) of theorem 2.3. Secondly, the estimation of each iteration with uniformly initial vector x^0 is very smooth, which results the parameter β_k goes to zero fast by the decrease condition of the TV function in formulae (*) and (★). On the other hand, the estimation by the TV-superiorized algorithm is very rough, which results the decrease of the TV function for relatively large

Table 3: TV, wavelet l^1 -norm, RMSE and iteration of the images in figure 5.

image	EM	TV-alg1	hard-alg1	soft-alg1
TV($\times 10^3$)	40.591	29.804	40.364	34.214
$l^1(\times 10^3)$	19.098	14.632	18.907	16.138
RMSE	0.2538	0.2112	0.2529	0.2266
iterations	13	13	13	13
image	-	TV-alg2	hard-alg2	soft-alg2
TV($\times 10^3$)	-	17.133	16.203	16.372
$l^1(\times 10^3)$	-	9.403	3.417	3.207
RMSE	-	0.1469	0.1683	0.1873
iterations	-	19	30	23

parameter β_k . Thirdly, the threshold techniques(hard and soft) always decrease the l^1 -norm, which results the conditions (*) and (★) for the l^1 -superiorized algorithm 2 become to only one condition about the decrease of K-L distance.

Experiment 4: modification of the TV-superiorized algorithm 1 and 2

This experiment is based on the data set 1. We modify the the TV-superiorized algorithms by discarding the decreasing condition of TV in formulae (*) and (★).

Figure 9 displays the reconstructed images by TV-superiorized EM algorithm 1 and 2 in absence of the decreasing condition of TV function. The quantitative values are tabulated in table 4. And figure 10 plots the variances of β_k versus iteration number.

As expectation, the parameter β_k converged to zero at much slower rate comparing with the TV-superiorized algorithm 2, which is similar to the l^1 -superiorized EM algorithm 2. However, this modification have very little effect on the TV-superiorized algorithm 1. The observations above show that the decreasing condition of TV is dominated for the TV-superiorized EM algorithm 2, while the condition 3 of theorem 2.3 for the TV-superiorized EM algorithm 1.

It is amazing that the modified algorithm also decrease the TV function, even the reconstructed image is better than that the TV-superiorized EM algorithm 2 , though we do not validate the decreasing condition of it.

Table 4: TV, wavelet l^1 -norm, RMSE and iteration of the images in figure 9 ($\times 10^3$)

image	image 1	image 2	image 3	image4
TV($\times 10^3$)	25.736	14.101	13.105	7.758
$l^1(\times 10^3)$	12.534	7.494	6.4517	3.6920
RMSE	0.1884	0.1278	0.2778	0.2075
iteration	13	27	8	16

Remark 4.1 *The experiments show that the convergence of the superiorized EM algorithm 2, though we can not prove it theoretically so far.*

Remark 4.2 *The parameter β_k represents the strength regularization of ϕ , so we can obtain the regularization reconstruction by terminating the iteration as long as β_k is smaller than a predefined number ϵ . The discussion about the selection of ϵ will be studied in future work.*

5 Conclusions and Discussions

In this paper, we discussed the convergence of the EM algorithm in the presence of perturbations, and proposed the superiorized EM algorithm 1 based on the convergent conditions. The modified version of algorithm 1 was proposed by relaxing the condition 3 of theorem 2.3. We gave the methods to generate the desired perturbation for TV- and l^1 -superiorization algorithms. The numerical experiments for SPECT reconstruction show that the superiorization algorithms is more stable and robust than the classic EM algorithm for low-count data set and initial point, and they could efficiently decrease the corresponding objective functions as expectation.

A lot work should be developed to about the superiorized EM iteration. Could we prove the convergence of the superiorized EM algorithm 2 theoretically? Secondly, we have not exact theories to support the amazing observation of the experiment 4. The observations of the numerical experiments, especially experiment 3 and

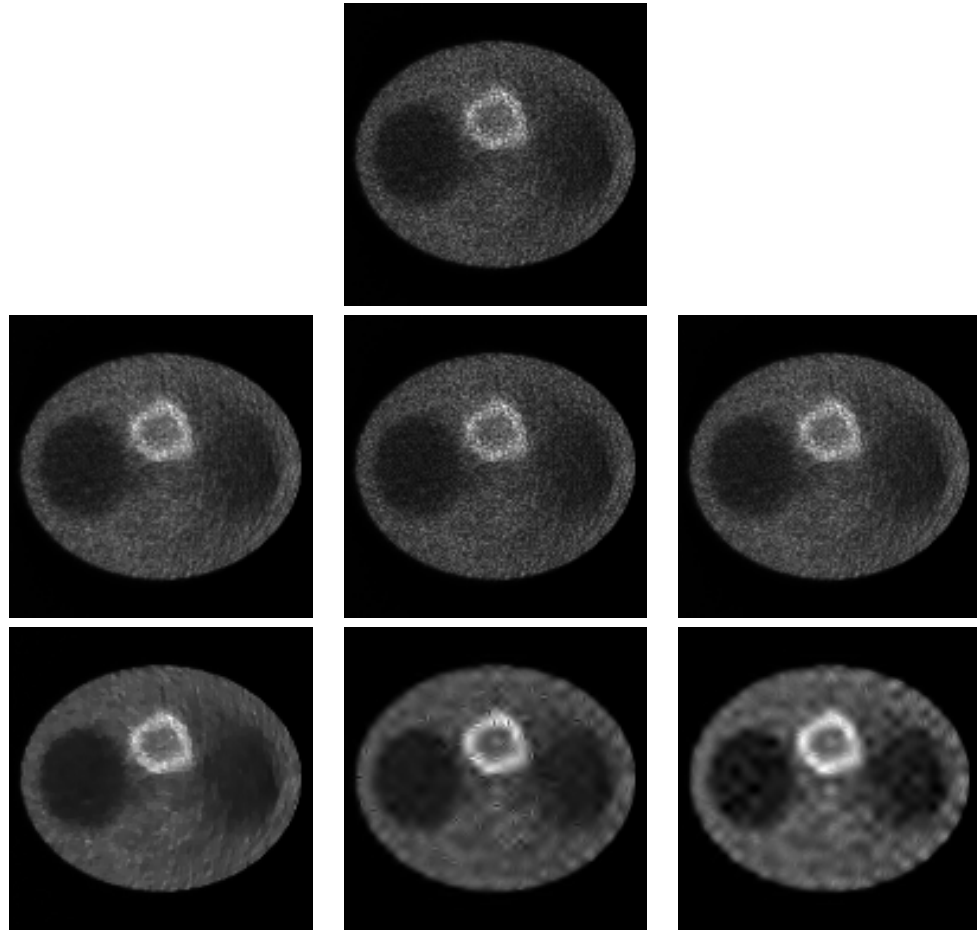


Figure 7: Images for experiment 3. The top image is reconstructed by the classic EM algorithm. The images in the column 1, 2 and 3 and the middle and bottom rows are reconstructed by the TV-, hard- and soft-superiorized EM algorithm 1 and 2, respectively.

4, would enlighten us to modify the the superiorized of the EM iteration.

In addition, the concept of superiorization is an emerging technique, and a lot of works need to be done and developed. For instance, we could not prove $\phi(x^*) \leq \phi(\hat{x})$ theoretically, where \hat{x} and x^* are the solutions by the classic iteration algorithm and the ϕ -superiorization version [18]. Secondly, the superiorization theories are mainly about algebraic iteration [18, 20] so far, and it is difficult to generalize the theory about algebraic iteration to other iteration algorithm. Specially, due to the nonlinearity of EM algorithm, we can not prove the bounded perturbation resilience(BPR) of EM algorithm, and the convergent conditions in this paper for the perturbed EM algorithm are much stronger than that for the perturbed algebra iteration.

Acknowledgement

This work is supported by the National Basic Research Program of China (2011CB809105) and NSF grants of China (61121002, 10990013). The authors are grateful for the helpful discussions with Professor Haomin Zhou(School of Mathematics, Georgia Institute of Technology).

References

- [1] M. N. Wernick and J. N. Aarsvold, *Emission Tomography: the Fundamentals of PET and SPECT*. California: Elsevier Academic Press, 2004.
- [2] G. T. Herman, *Image Reconstruction from Projection: the Fundamentals of Computerized Tomography*. New York: Academic Press, 1980.

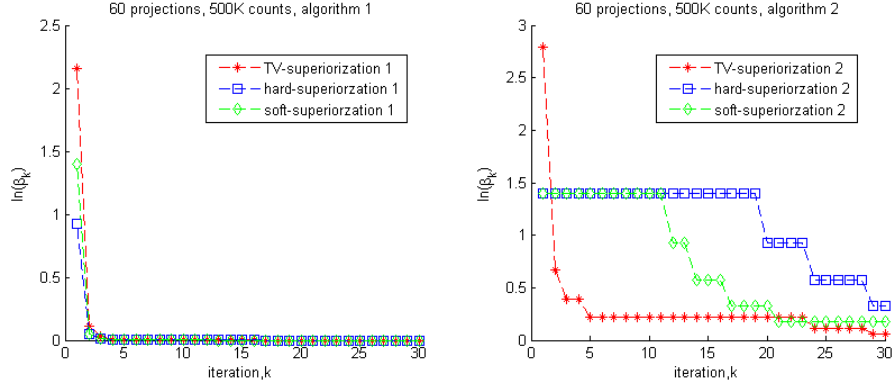


Figure 8: the variance plot of K-distance, MSE and β_k versus iteration for data set 1 and initial point with random values.

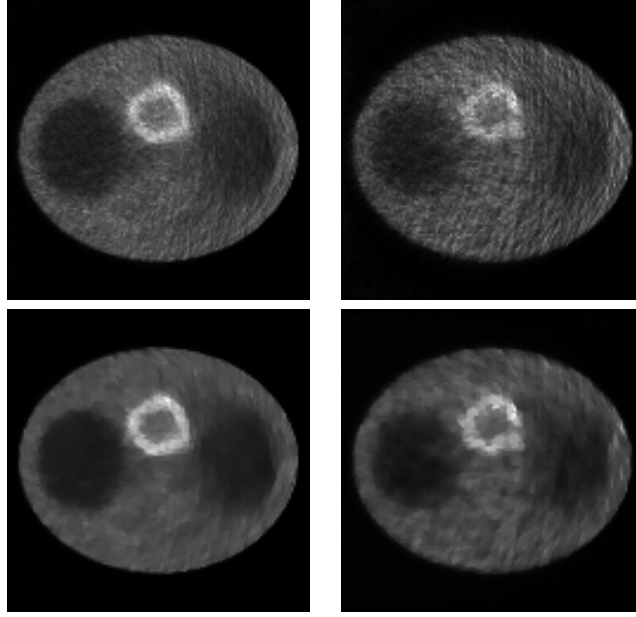


Figure 9: Images of experiment 4. The images in left and right rows are reconstructed from data set 1 and 2, while the images in first and second rows are reconstructed by modified versions of TV-superiorized EM algorithm 1 and 2, respectively. From top to bottom and left to right, the images are labeled image 1 to 4 for reference.

- [3] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*. Philadelphia: Society for Industrial and Applied Mathematics, 2001.
- [4] O. Tretiak and C. Metz, “The exponential radon transform,” *SIAM Journal on Applied Mathematics*, vol. 39, pp. 341–354, 1980.
- [5] F. Natterer, “Inversion of the attenuated radon transform,” *Inverse Problems*, vol. 17, pp. 113–119, 2001.
- [6] R. G. Novikov, “An inversion formula for the attenuated x-ray transformation,” vol. 40, pp. 145–167, 2002.
- [7] F. Noo and J.-M. Wagner, “Image reconstruction in 2d spect with 180° acquisition,” *Inverse Problems*, vol. 17, pp. 1357–71, 2001.
- [8] F. Noo, M. Defrise, J. D. Pack, and R. Clackdoyle, “Image reconstruction from truncated data in single-photon emission computed tomography with uniform attenuation,” *Inverse Problems*, vol. 23, no. 2, pp. 645–667, 2007.
- [9] L. A. Shepp and Y. Vardi, “Maximum likelihood restoration for emission tomography,” *IEEE Transaction Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

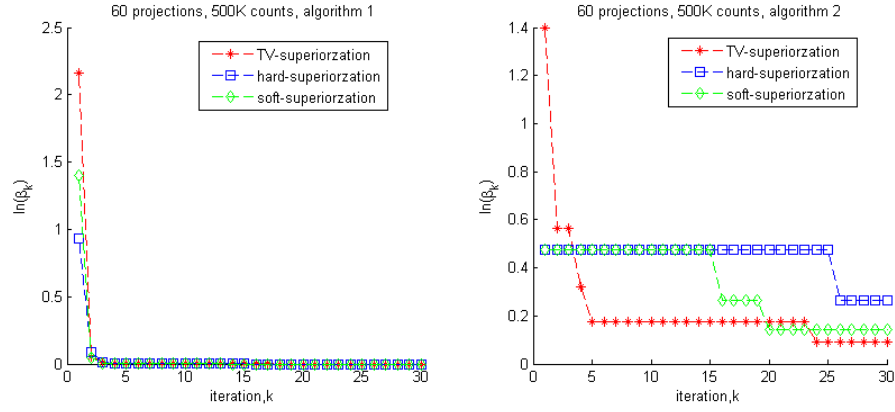


Figure 10: The plot of β_k variance versus iteration k for modified TV-superiorized EM algorithm 1(left) and 2(right).

- [10] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transaction Medical Imaging*, vol. 13, no. 4, pp. 601–609, 1994.
- [11] I.-T. Hsiao and H.-M. Huang, "An accelerated ordered subsets reconstruction algorithm using an accelerating power factor for emission tomography," *Physics in Medicine and Biology*, vol. 55, pp. 599–614, 2010.
- [12] Y. Vardi and L. K. L. A. Shepp, "A statistical model for positron emission tomography," *Journal of the American Statistical Association*, vol. 80, pp. 8–20, 1985.
- [13] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography," *Journal of Theoretical Biology*, vol. 29, pp. 471–482.
- [14] Y. Censor, T. Elfving, and G. T. Herman, "Averaging strings of sequential iterations for convex feasibility problems," in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, pp. 101–113, 2001.
- [15] R. Davidi, G. T. Herman, and Y. Censor, "Perturbation-resilient block-iterative projection methods with application to image reconstruction from projection," *International Transaction in Operational Research*, pp. 505–524, 2009.
- [16] D. Butnariu, R. Davidi, G. T. Herman, and I. G. Kazantsev, "Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problem," *IEEE Journal of Selected Topics Signal Processing*, vol. 1, no. 4, pp. 540–547, 2007.
- [17] G. T. Herman and R. Davidi, "Image reconstruction from a small number of projections," *Inverse Problems*, vol. 24, no. 4, p. 045011, 2008.
- [18] Y. Censor, R. Davidi, and G. T. Herman, "Perturbation resilience and superiorization of iteration algorithm," *Inverse Problems*, vol. 26, pp. 1–12, 2010.
- [19] E. Garduo, G. T. Herman, and R. Davidi, "Reconstruction from a few projections by l^1 minimization of the haar transform," *Inverse Problems*, vol. 24, no. 4, p. 0550061, 2011.
- [20] T. Nikazad, R. Davidi, and G. T. Herman, "Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction," *Inverse Problems*, vol. 28, no. 3, p. 035005, 2012.
- [21] W. Jin, Y. Censor, and M. Jiang, "A heuristic superiorization-like approach to bioluminescence tomography," in *Proceedings of the International Federation for Medical and Biological Engineering (IFMBE)*, Springer-Verlag, to appear, 2012.
- [22] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259–268, 1992.
- [23] E. Candes, J. Romberg, and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n ," *Arxiv preprint math*, 2005.

- [24] D. L. Snyder, T. J. Schulz, and J. A. ÓSullivan, “Deblurring subject to nonnegativity constraints,” *IEEE Trans. Signal Processing*, vol. 40, pp. 1143–1150, 1992.
- [25] D. L. Snyder, T. J. Schulz, and J. A. O. Sullivan, “Deblurring subject to nonnegativity constraints,” *IEEE Transaction on Signal Processing*, vol. 40, no. 5, pp. 1143–1150, 1992.
- [26] E. Y. Sidky, C.-M. Kao, and X. Pan, “Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct,” *X-Ray Science and Technology*, vol. 14, p. 119C139, 2006.
- [27] S. J. LaRoque, E. Y. Sidky, and X. Pan, “Accurate image reconstruction from few-view and limited-angle data in diffraction tomography,” *Optical Society*, vol. 25, no. 7, pp. 1172–1872, 2008.
- [28] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,”
- [29] H. Yu and G. Wang, “Compressed sensing based interior tomography,” *Physical Medicine Biology*, vol. 54, pp. 2791–805, 2009.
- [30] Y. Censor, D. E. Gustafson, A. Lent, and H. Tuy, “A new approach on the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients,” *IEEE Transactions on Nuclear Science*, vol. 26, pp. 2775–2779, 1979.
- [31] J. A. fessler, “Matlab code for emission tomography.” <http://www.eecs.umich.edu/fessler/>.